



The overlap gap property: A topological barrier to optimizing over random structures

David Gamarnik^{a,b,1}

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved August 19, 2021 (received for review May 5, 2021)

The problem of optimizing over random structures emerges in many areas of science and engineering, ranging from statistical physics to machine learning and artificial intelligence. For many such structures, finding optimal solutions by means of fast algorithms is not known and often is believed not to be possible. At the same time, the formal hardness of these problems in the form of the complexity-theoretic NP -hardness is lacking. A new approach for algorithmic intractability in random structures is described in this article, which is based on the topological disconnectivity property of the set of pairwise distances of near-optimal solutions, called the Overlap Gap Property. The article demonstrates how this property 1) emerges in most models known to exhibit an apparent algorithmic hardness; 2) is consistent with the hardness/tractability phase transition for many models analyzed to the day; and, importantly, 3) allows to mathematically rigorously rule out a large class of algorithms as potential contenders, specifically the algorithms that exhibit the input stability (insensitivity).

algorithms and computation | phase transition | random structures | spin glasses

Optimization problems involving uncertainty emerge in many areas of science and engineering, including statistics, machine learning and artificial intelligence, computer science, physics, biology, management science, economics, and social sciences. The exact nature and the sources of uncertainty vary from field to field. The modern paradigm of Big Data brought forward optimization problems involving, in particular, many dimensions, thus creating a new metafield of “high-dimensional probability” and “high-dimensional statistics” (1–3). Two recent special semester programs at the Simons Institute for the Theory of Computing were devoted to this burgeoning topic,* among a multitude of other conferences and workshops. While many of the optimization problems involving randomness can be solved to optimality by fast, and in fact, sometimes borderline trivial, algorithms, other problems have resisted decades of attempts, and slowly, it has been accepted that these problems are likely nonamenable to fast algorithms, with polynomial time algorithms broadly considered to be the gold standards for what constitutes fast algorithms. The debate surrounding the actual algorithmic hardness of these problems, though, is by no means

settled, unlike its “worst-case” algorithmic complexity-theoretic counterparts, expressed in the form of the widely believed $P \neq NP$ conjecture.

What is the “right” algorithmic complexity theory explaining the persistent failure to find tractable algorithms for these problems? We discuss in this paper some existing theories and explain their shortcomings in light of what we know now about the state-of-the-art algorithms. We then propose an approach, largely inspired by the field of statistical physics, which offers a new topological/geometric theory of algorithmic hardness that is based on the disconnectivity of the overlaps of near-optimal solutions, dubbed the “Overlap Gap Property” (OGP). The property can and has been rigorously verified for many concrete models and, importantly, can be used to mathematically rule out large classes of algorithms as potential contenders, specifically algorithms exhibiting a form of input stability. This includes both classical and quantum (Quantum Approximate Optimization Algorithm [QAOA]) algorithms (4, 5), which recently gained a lot of attention as one of the most realistic algorithms to be implementable on quantum computers (6). A widely studied random Number

^aOperations Research Center, Massachusetts Institute of Technology, Cambridge, MA 02140; and ^bSloan School of Management, Massachusetts Institute of Technology, Cambridge, MA 02140

Author contributions: D.G. designed research, performed research, contributed new reagents/analytic tools, and wrote the paper.

The author declares no competing interest.

This article is a PNAS Direct Submission.

Published under the [PNAS license](#).

¹Email: Gamarnik@mit.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2108492118/-/DCSupplemental>.

Published October 1, 2021.

*<https://simons.berkeley.edu/programs/hd20>; <https://simons.berkeley.edu/programs/si2021>.

Partitioning problem will be used to illustrate both the OGP and the mechanics by which this property manifests itself as a barrier for tractable algorithms.

The Largest Clique of a Random Graph: The Most “Embarrassing” Open Problem in Random Structures

Imagine a club with N members, in which about 50% of the $N(N-1)/2$ member pairs know each other personally, and the remaining 50% of the members do not. You want to find a largest clique in this club, namely, the largest group of members out of the N members who all know each other. What is the typical size c^* of such a clique? How easy is it to find one? This question can be modeled as the problem of finding a largest fully connected subgraph (formally called a “clique”) in a random Erdős–Rényi graph $\mathbb{G}(N, 1/2)$, which is a graph on N nodes, where every pair of nodes is connected with probability $1/2$, independently for all pairs. The first question regarding the largest clique size c^* is a textbook example of the so-called probabilistic method, a simple application of which tells us that c^* is likely to be near $2 \log_2 N$ with high degree of certainty as N gets large (7). A totally different matter is the problem of actually finding such a clique, and this is where the embarrassment begins. Richard Karp, one of the founding fathers of the algorithmic complexity theory, observed in his 1976 paper (8) that a very simple algorithm, both in terms of the analysis and the implementation, finds a clique of roughly half-optimum size, namely, with about $\log_2 N$ members, and challenged to do better. The problem is still open, and this is embarrassing for two reasons: 1) The best-known algorithm, namely, the one above, is also an extremely naive one. So it appears that the significant progress that the algorithmic community has achieved over the past decades in constructing ever more clever and powerful algorithms is totally helpless in improving upon this extremely simple and naive algorithm. 2) We don’t have a workable theory of algorithmic hardness, which rigorously explains why finding cliques of size half-optimum is easy, and improving on it does not seem possible within the realm of polynomial time algorithms. The classical $P \neq NP$ paradigm and its variants are of no help here. A variant of this problem, called the Hidden Clique Problem, was introduced in a seminal paper by Jerrum (9), and it exhibits a similar algorithmic gap.

The largest clique problem turns out to be one of very many other problems exhibiting a similar phenomenon: Using nonconstructive analysis method, one shows that the optimal value of some optimization problem involving randomness is some value c^* , the best known polynomial time algorithm achieves value c_{ALG} , and there is a significant gap between the two: $c_{\text{ALG}} < c^*$. A partial list (but one growing seemingly at the speed of the Moore’s Law) is the following: the randomly generated constraint satisfaction problem (also known as [aka] the random K -satisfiability [K-SAT] problem), the largest independent set in a random graph, proper coloring of a random graph, finding a ground state of a spin-glass model in statistical mechanics, discovering communities in a network (the so-called community detection problem), group testing, statistical learning of a mixture of Gaussian distributions, the sparse linear regression problem, the sparse covariance estimation, the graph matching problem, the spiked tensor problem, the Number Partitioning problem, and many other problems.

The Number Partitioning problem is motivated in particular by the statistical problem of designing a randomized controlled study with two groups possessing roughly “equal” attributes. It is also a special case of the bin-packing problem and also has been

widely studied in the statistical physics literature. The Number Partitioning problem will serve as one of the running examples in this article, so we introduce it now. Given N items with weights X_1, \dots, X_N , the goal is to split it into two groups so that the difference of total weights in two groups is the smallest possible. Namely, the goal is finding a subset $\{I \subset \{1, \dots, N\}\}$ such that $|\sum_{i \in I} X_i - \sum_{i \notin I} X_i|$ is as small as possible. An NP-hard problem in the worst case (10), it is more tractable in the presence of randomness. Suppose the weights X_i are generated independently, according to the standard Gaussian distribution $\mathcal{N}(0,1)$. A rather straightforward application of the same probabilistic method shows that the optimum value is typically $\sqrt{N}2^{-N}$ for large N . Namely, the value of c^* in our context is $\sqrt{N}2^{-N}$. An algorithm proposed by Kamarkar and Karp in 1982 (11) achieves the value of order $c_{\text{ALG}} = \sqrt{N}e^{-c \log_2^2 N}$, with value c predicted (though not proven) to be $1/(2 \log 2) \approx 0.721..$ (12). The multiplicative gap between the two is thus very significant (exponential): $c_{\text{ALG}}/c^* = 2^{N-O(\log_2^2 N)}$. No improvement of this result is known to date, and no algorithmic hardness result is known either.

In Search of the Right Algorithmic Complexity Theory

We now describe some existing approaches for understanding the algorithmic complexity and discuss their shortcomings in explaining the existential/algorithmic gap exhibited by the multitude of problems described above. Starting with the classical algorithmic complexity theory based on the algorithmic complexity classes such as P , NP , etc., this theory is of no help, since these complexity classes are based on the worst-case assumptions on the problem description. For example, assuming the widely believed conjecture that $P \neq NP$, finding a largest clique in a graph with N nodes within a multiplicative factor $N^{1-\delta}$ is not possible by polynomial time algorithms (13) for any constant $\delta \in (0,1)$ if $P \neq NP$. This is a statement, however, about the algorithmic problem of finding large cliques in all graphs, and it is in sharp contrast with the factor $1/2$ achievable by polynomial time algorithms for random graphs $\mathbb{G}(N, 1/2)$, according to the discussion above.

A significant part of the algorithmic complexity theory does in fact deal with problems with random input, and here, impressive random-case to worst-case reductions are achievable for some of the problems. These problems enjoy a wide range applications in cryptography, where the average case-hardness property is paramount for designing secure cryptographic protocols. For example, if there exists a polynomial time algorithm for computing the permanent of a matrix with independent and identically distributed (i.i.d.) random inputs, then, rather surprisingly, one can use it to design a polynomial time algorithm for all matrices (14). The latter problem is known to be in the $\#P$ complexity class, which subsumes NP . A similar reduction exists (15) for the problem of computing the partition function of a Sherrington–Kirkpatrick model described below, thus implying that computing partition functions for spin-glass models is not possible by polynomial time algorithms unless $P = \#P$. Another problem admitting average-case to worst-case reduction is the problem of finding a shortest vector in a lattice (16). The random to worst-case types of reduction described above would be ideal for our setting, as they would provide the most compelling evidence of hardness of these problems. For example, it would be ideal to show that finding a clique of size at least, say, $(3/2) \log_2 N$ in $\mathbb{G}(N, 1/2)$ implies a polynomial time algorithm with the same approximation guarantee for all graphs. Unfortunately, such a result appears out of reach for the existing proof techniques.

Closer to the approach described in this paper, and the one that in fact has largely inspired the present line of work, is a range of theories based on the solution space geometry of the underlying optimization problem. This approach emerged in the statistical physics literature, specifically, the study of spin glasses, and more recently found its way to questions above and beyond statistical physics, in particular, questions arising in the context of studying neural networks, as in ref. 17. The general philosophy of this take on the algorithmic complexity is that when the problem appears to be algorithmically hard, this somehow should be reflected in the nontrivial geometry of optimal or near-optimal solutions. One of the earliest such approaches was formulated for the decision problems, such as random constraint satisfaction problems (aka the random K-SAT problem). It links the algorithmic hardness with the proximity to the satisfiability phase-transition threshold (18, 19) and the order (first vs. second) of the associated phase transition (20). To elaborate, we need to introduce the random K-SAT problem first. It is a Boolean constraint satisfaction problem involving N variables x_1, \dots, x_N defined as a conjunction Φ of M clauses $C_1 \wedge C_2 \wedge \dots \wedge C_M$, where each clause C_j is a disjunction of exactly K variables from x_1, \dots, x_N or their negation. Thus, each C_j is of the form $x_{j,1} \vee x_{j,2} \vee \dots \vee x_{j,K}$. An example of such a formula with $N = 10$, $M = 4$ and $K = 3$ is, say, $(x_2 \vee x_5 \vee x_6) \wedge (\neg x_1 \vee \neg x_2 \vee x_9) \wedge (\neg x_6 \vee x_8 \vee \neg x_{10}) \wedge (\neg x_3 \vee x_2 \vee x_7)$. A formula is called satisfiable if there exists an assignment of Boolean variables x_i , $1 \leq i \leq N$ to values zero and one such that the value of the formula is one.

A random instance of a K-SAT problem is obtained by selecting variables into each clause C_j uniformly at random, independently for all j , and applying the negation operation \neg with probability $1/2$ independently for all variables. The random K-SAT problem is viewed by statistical physicists as a problem exhibiting the so-called frustration property, which is of great interest to physics of spin glasses, and, hence, it has enjoyed a great deal of attention in the statistical physics literature (21, 22).

As it turns out, for each K , there is a conjectured critical threshold $\alpha_{\text{SAT}} = M/N$ for the satisfiability property, which was rigorously proven (23) for large enough K . In particular, for every $\alpha < \alpha_{\text{SAT}}$, the formula admits a satisfying assignment when $M/N < \alpha$, and for every $\alpha > \alpha_{\text{SAT}}$, the formula does not admit a satisfying assignment when $M/N > \alpha$, both with overwhelming probability as $N \rightarrow \infty$. The sharpness of this transition was established rigorously earlier by general methods in ref. 24. The algorithmic counterparts, however, came short of achieving the value α_{SAT} for every $K \geq 3$ (more on this below). Even before the results in refs. 23 and 24, it was conjectured in the physics literature that perhaps the existence of the sharp phase-transition property itself is the culprit for the algorithmic hardness (19, 20). This was argued by studying the heuristic running times of finding the satisfying solutions or demonstrating nonexistence of ones for small values of K and observing that the running time explodes near α_{SAT} and subsides when M/N is far from α_{SAT} on either side. The values of α_{SAT} are known thanks to the powerful machinery of the replica symmetry methods developed in seminal works of Parisi (21, 25). These methods give very precise predictions for the values of α_{SAT} for every K . For example, α_{SAT} is ~ 4.26 when $K = 3$. An in-depth analysis of the exponent of the running times was reported in ref. 19.

The theory postulating that the algorithmic hardness of the K-SAT problem is linked with the satisfiability phase transition, however, appears inconsistent with the later rigorous discoveries obtained specifically for large values of K . In particular, while α_{SAT}

is known to be approximately $2^K \log 2$ for large values of K , all of the known algorithms stall long before that, and specifically at $\alpha_{\text{ALG}} = (2^K/K) \log K$ (26). Furthermore, there is evidence that breaking this barrier might be extremely challenging. This was argued by proving that the model undergoes the so-called *clustering* phase transition near α_{ALG} (27, 28) (more on this below), and also by ruling out various families of algorithms. These algorithms include sequential local algorithms and algorithms based on low-degree polynomials (using the OGP discussed in the next section) (29, 30), the Survey Propagation algorithm (31), and a variant of a random-search algorithm called WalkSAT (32). The Survey Propagation algorithm is a highly effective heuristics for finding satisfying assignments in random K-SAT and many other similar problems. It is particularly effective for low values of K and beats many other existing approaches in terms of the running times and sizes of instances it is capable to handle (33) (see also ref. 34 for the perspective on the approach). It is entirely possible that this algorithm and even its more elementary version, the Belief Propagation guided decimation algorithm, solves the satisfiability problem for small values of K . An evidence of this can be found in ref. 35, and rigorous analysis of this algorithm was conducted by Coja-Oghlan in ref. 36, which shows it to be effective when $\alpha \leq 2^K/K$. However, the analysis by Hetterich in ref. 31 rules out the Survey Propagation algorithm for sufficiently large values of K , beyond the α_{ALG} threshold, which we recall is $(2^K/K) \log K$. Additionally, the theory of algorithmic hardness based on the existence of the satisfiability-type phase transition does not appear to explain the algorithmic hardness associated with *optimization* problems, such as the largest clique problem in $\mathbb{G}(N, 1/2)$.

The clustering property was rigorously established (27, 28) for random K-SAT for values of α above α_{Clust} , which is known to be close to α_{ALG} for large K . We specifically refer to this as *weak clustering* property in order to distinguish it from the *strong clustering* property, both of which we now define. (While the distinction between weak and strong clustering was discussed in many prior works, the choice of the terminology is entirely by the author.) The model exhibits the weak clustering property if there exists a subset Σ of satisfying assignments, which contains all but exponentially small in N fraction of the satisfying assignments and which can be partitioned into subsets (clusters) separated by order $O(N)$ distances, such that within each cluster, one can move between any two assignments by changing, at most, constantly many $O(1)$ bits. In other words, by declaring two assignments σ and τ connected if τ can be obtained from σ by flipping the values of, at most, $O(1)$ variables, the set Σ consists of several disconnected components separated by linear in N distances. This is illustrated in *SI Appendix, Fig. S1*, where blue regions represent clusters in Σ . The gray "tunnel"-like sets depicted in the figure are part of the complement (exception) subset Σ^c of the set of satisfying assignments, which may potentially connect (tunnel between) the clusters. Furthermore, the clusters are separated by exponentially large in N cost barriers, meaning that any path connecting two solutions from different clusters (which then necessarily contains assignments violating at least some of the constraints), at some point, contains in it assignments, which, in fact, violate order $O(N)$ constraint. For this reason, the clustering property is sometimes referred to as an "energetic" barrier.

As mentioned earlier, for large values of K , the onset of weak clustering occurs near $2^K \log 2/K$, and, in fact, the suggestion that this value indeed corresponds to the algorithmic threshold α_{ALG} was partially motivated by this discovery of weak clustering property

around this value. See also ref. 37 for the issues connecting the clustering property with the algorithmic hardness. Below α_{ALG} , the bulk of the set of satisfying assignments constitutes one connected subset.

On the other hand, the *strong* clustering property is the property that all satisfying assignments can be partitioned into clusters like the ones above, however, with no exceptions. Namely, Σ is the set of all satisfying assignments. This can be visualized from *SI Appendix, Fig. S1*, where the gray tunnels would be removed. It turns out that for large values of K , the model does exhibit the strong clustering property as well, but the known lower bounds for it are of the order 2^K , as opposed to $2^K/K$, for large K . While not proven formally, it is expected though that the onset of the strong clustering property indeed occurs at values order 2^K , as opposed to $2^K/K$, for large K . Namely, the weak and strong clustering properties appear at different in K scales.

The existence of the strong clustering property is established as an implication of the OGP (which is the subject of this paper), and this was the approach used in refs. 27 and 28. In contrast, the weak clustering property is established by using the so-called *planting* solution, which amounts to consider the solution-space geometry from the perspective of a uniformly at random selected satisfying assignment.

The necessity of distinguishing between the two modes of clustering described above is driven by algorithmic implications. Solutions generated by most algorithms typically are not generated uniformly at random and, thus, in principle, can easily fall into the exception set Σ^c . Thus, the obstruction arguments based somehow on linear separations between the clusters and exponentially large energetic barriers might simply be of no algorithmic relevance; see the later *OGP, the Clustering Property, and the Curious Case of the Perceptron Model* section devoted to the relationship between clustering and the OGP.

As the notion of “exception” size in the definition of weak clustering property hinges on the counting measure associated with the space of satisfying assignments (and the uniform measure was assumed in the discussion above), a potential improvement might stem from changing this counting measures and introducing solution-specific weights. Interestingly, for small values of K , this can be done effectively delaying the onset of weak clustering, as was shown by Budzynski et al. in ref. 38. Unfortunately, though, for large K , the gain is only in second-order terms, and up to those orders, the threshold for the onset of weak clustering based on biased counting measures is still $2^K \log 2/K$, as shown by Budzynski and Semerjian in ref. 39. This, arguably, provides an even more robust evidence that this value marks a phase-transition point of fundamental nature.

More recently, algorithmic barriers were suggested to be linked with refined properties of the clustering phase transition, specifically, the so-called condensation-phase transition (40) and the emergence of so-called *frozen* variables. Algorithmic implications of both are discussed in ref. 41. Using again a powerful machinery of the replica symmetry breaking, providing non-rigorous, though highly believable, predictions, one identifies another phase transition α_{COND} satisfying $\alpha_{\text{ALG}} < \alpha_{\text{COND}} < \alpha_{\text{SAT}}$ for large K . As the value M/N passes through α_{COND} , the number of solution clusters covering the majority of satisfying assignment in the random K-SAT problem drops dramatically from the exponentially many to only constantly many ones, with the largest cluster containing a nontrivial fraction of all of the assignments (40) (*SI Appendix, Fig. S1*). At this stage, a long-range dependence

between the spin magnetization (appropriately defined) emerges, and, furthermore, at this stage, the random overlaps (inner products) of pairs of assignments generated according to the associated Gibbs measure have a nontrivial limiting distribution, described by the so-called Parisi measure.

Even before the value M/N exceeds the value α_{COND} for large K , each cluster contains a nontrivial fraction of the frozen variables, which are variables always taking the same values within a given cluster of satisfying assignment. The authors in ref. 41 conjecture that the emergence of frozen variables is the primary culprit of the algorithmic hardness and construct a variant of the Survey Propagation algorithm called the Backtracking Survey Propagation algorithm, which is conjectured to be effective all the way up to $M/N < \alpha_{\text{COND}}$. Numerical simulations demonstrate an excellent performance of this algorithm on randomly generated instances of K-SAT for small values of K .

The conjecture linking the emergence of frozen variables with algorithmic hardness stumbles, however, upon similar challenges as the theory based on the existence of the phase transition. A rigorous evidence has been established that when K is large, α_{COND} is of the order $2^K \log 2 - C$ for some (explicitly computable) value C , which does not depend on K . In particular, α_{COND} is substantially above the value α_{ALG} at which all known algorithms fail, and, furthermore, classes of algorithms, including the standard Survey Propagation algorithm, are ruled out, as discussed above. Of course, the nonexistence of evidence is not necessarily the evidence of nonexistence, and it is very much possible that in the future, successful polynomial time algorithms will be constructed in the regime $\alpha_{\text{ALG}} < M/N < \alpha_{\text{COND}}$. But there is another limitation of the theory of algorithmic hardness based on the emergence of frozen variables: Just like the notion of the satisfiability phase transition, the notion of frozen variables does not apply to problems of optimization type, such as the largest clique problem, or an optimization problem appearing in the context of spin-glass models—for example, the problem of finding a ground state of a spherical p -spin-glass model. In this model, spins take values in a continuous range, and, thus, the notion of frozenness, which is exclusively discrete, is lost.

Topological Complexity Barriers: The OGP

We now introduce a new approach for predicting and proving algorithmic complexity for solving random constraint satisfaction problems and optimization problems on random structures, such as the ones introduced in sections *The Largest Clique of a Random Graph: the Most “Embarrassing” Open Problem in Random Structures* and *In Search of the Right Algorithmic Complexity Theory*. The approach bears a lot of similarity with the predictions based on the clustering property described in the latter of the two aforementioned sections, but has important and nuanced differences, which allow us to rule out large classes of algorithms, something that did not seem possible before. An important special case of such algorithms is stable algorithms, namely, algorithms exhibiting low sensitivity to the data input. A special, yet very broad and powerful, class of such algorithms is algorithms based on low-degree polynomials. A fairly recent stream of research (42–44) puts forward an evidence that, in fact, algorithms based on low-degree polynomials might be the most powerful class of polynomial time algorithms for optimization problems on random structures, such as the problems discussed in this paper.

A Generic Formulation of the Optimization Problem. To set the stage, we consider a generic optimization problem $\min_{\sigma} \mathcal{L}(\sigma, \Xi_N)$. Here, \mathcal{L} encodes the objective function (cost) to be minimized. The solutions σ lie in some ambient solution space Σ_N , which is typically very high-dimensional, often discrete, with N encoding its dimension. The space Σ_N is equipped with some metric (distance) $\rho_N(\sigma, \tau)$ defined for each pair of solutions $\sigma, \tau \in \Sigma_N$. For the max-clique problem, we can take $\Sigma_N = \{0, 1\}^N$. For the Number Partitioning problem, we can take $\Sigma_N = \{-1, 1\}^N$. Ξ_N is intended to denote the associated randomness of the problem. So, for example, for the max-clique problem, Ξ_N encodes the random graph $\mathbb{G}(N, 1/2)$, and $\sigma \in \{0, 1\}^N$ encodes a set of nodes constituting a clique, with $\sigma_i = 1$ if node i is in the clique and $=0$ otherwise. We denote by ξ an instance generated according to the probability law of Ξ_N . In the present context, ξ is any graph, for which the likelihood of generating is $2^{-\frac{N(N-1)}{2}}$. Finally, $-\mathcal{L}(\sigma, \Xi_N)$ is the number of ones in the vector σ . Now, if σ_N does not actually encode a clique (that is, for at least one of the edges (i, j) of the graph, we have $\sigma_{N,i} = \sigma_{N,j} = 1$), we can easily augment the setup by declaring $\mathcal{L}(\sigma, \Xi_N) = \infty$ for such "nonclique" encoding vectors. For the Number Partitioning problem, Ξ_N is just the probability distribution associated with an N -vector of independent standard Gaussians. For each $\sigma \in \Sigma_N = \{-1, 1\}^N$ and each instance ξ of Ξ_N , the value of the associated partition is $\mathcal{L}(\sigma, \xi) = |\langle \sigma, \xi \rangle|$. Here, $\langle x, y \rangle$ denotes an inner product between vectors x and y . Thus, for the largest clique problem, we have that the random variable $c^* = -\min_{\sigma \in \Sigma_N} \mathcal{L}(\sigma, \Xi_N)$ is approximately $2 \log_2 N$ with high degree of likelihood. For the Number Partitioning problem, we have that $c^* = \min_{\sigma \in \Sigma_N} \mathcal{L}(\sigma, \Xi_N)$ is approximately $\sqrt{N} 2^{-N}$ with high degree of likelihood as well.

The OGP and its Variants. We now introduce the OGP and its several variants. The term was introduced in ref. 45, but the property itself was discovered by Achlioptas and Ricci-Tersenghi (27) and Mezard, Mora, and Zecchina (28). The definition of the OGP pertains to a particular instance ξ of the randomness Ξ_N . We say that the optimization problem $\min_{\sigma} \mathcal{L}(\sigma, \xi)$ exhibits the OGP with values $\mu > 0, 0 \leq \nu_1 < \nu_2$ if for every two solutions σ, τ , which are μ -optimal in the additive sense, namely, satisfy $\mathcal{L}(\sigma, \xi) \leq c^* + \mu, \mathcal{L}(\tau, \xi) \leq c^* + \mu$, it is the case that either $\rho_N(\sigma, \tau) \leq \nu_1$ or $\rho_N(\sigma, \tau) \geq \nu_2$. Intuitively, the definition says that every two solutions that are "close" (within an additive error μ) to optimality are either close (at most distance ν_1) to each other or "far" (at least distance ν_2) from each other, thus exhibiting a fundamental topological discontinuity of the set of distances of near-optimal solutions. In the case of random instances Ξ_N , we say that the problem exhibits the OGP if the problem $\min_{\sigma} \mathcal{L}(\sigma, \xi)$ exhibits the OGP with high likelihood when ξ is generated according to the law Ξ_N . An illustration of the OGP is depicted in *SI Appendix, Fig. S2*. The notion of the "overlap" refers to the fact that distances in normed space are directly relatable to inner products, commonly called *overlaps* in the spin-glass theory, via $\|\sigma - \tau\|_2^2 = \|\sigma\|_2^2 + \|\tau\|_2^2 - 2\langle \sigma, \tau \rangle$ and the fact that solutions σ, τ themselves often have identical or close to identical norms. The OGP is of interest only for certain choices of parameters μ, ν_1, ν_2 , which are always problem-dependent. Furthermore, all of these parameters along with the optimal value c^* are usually dependent on the problem size, such as the number of Boolean variables in the K-SAT model or the number of nodes in a graph. In particular, the property is of interest only if there are pairs of μ -optimal solutions σ, τ satisfying both $\rho_N(\sigma, \tau) \leq \nu_1$ and $\rho_N(\sigma, \tau) \geq \nu_2$ property. The first

case is trivial, as we can take $\tau = \sigma$, but the existence of pairs with distance at least ν_2 needs to be verified. Establishing the OGP rigorously can be either straightforward or technically very involved, again depending on the problem. It should not be surprising that the presence of the OGP presents a potential difficulty in finding an optimal solution, due to the presence of multiple local minima, similarly to the lack of convexity property. An important distinction should be noted, however, between the OGP and the lack of convexity property. The function $\mathcal{L}(\sigma, \xi)$ can be nonconvex, but not exhibiting the OGP, as depicted in *SI Appendix, Fig. S3*. Thus, the "common" intractability obstruction presented by nonconvexity is not identical with the OGP. Also, the solution space Σ_N is often discrete (such as the binary cube), rendering the notions of convexity nonapplicable.

We now extend the OGP definition to the so-called *ensemble* OGP, abbreviated as the e-OGP, and the *multi*-OGP, abbreviated as the m-OGP. We say that a set of problem instances Ξ satisfies the e-OGP with parameters $\mu > 0, 0 \leq \nu_1 < \nu_2$ if for every pair of instances $\xi, \psi \in \Xi$, for every μ -optimal solution σ of the ξ instance (namely, $\mathcal{L}(\sigma, \xi) \leq \min_{\sigma'} \mathcal{L}(\sigma', \xi) + \mu$), and every μ -optimal solution τ of the ψ instance (namely, $\mathcal{L}(\tau, \psi) \leq \min_{\tau'} \mathcal{L}(\tau', \psi) + \mu$), it is the case that either $\rho_N(\sigma, \tau) \leq \nu_1$ or $\rho_N(\sigma, \tau) \geq \nu_2$, and in the case when instances ξ and ψ are probabilistically independent, the former case (i.e., $\rho_N(\sigma, \tau) \leq \nu_1$) is not possible. The set Ξ represents a collection of problem instances over which the optimization problem is considered. For example, Ξ might represent a collection of correlated random graphs $\mathbb{G}(N, 1/2)$ or random matrices or random tensors. Indeed, below, we will provide an example of a family of correlated random graphs $\mathbb{G}(N, 1/2)$ as a main example of the case when the e-OGP holds. We see that the OGP is the special case of the e-OGP when Ξ consists of a single instance ξ .

Finally, we say that a family of instances Ξ satisfies the m-OGP with parameters $\mu, \nu_1 < \nu_2$ and m if for every m -collection of instances $\xi_1, \dots, \xi_m \in \Xi$ and every collection of solutions $\sigma_1, \dots, \sigma_m$ that are μ -optimal with respect to ξ_1, \dots, ξ_m , at least one pair σ_i, σ_j satisfies $\rho_N(\sigma_i, \sigma_j) \leq \nu_1$ or $\rho_N(\sigma_i, \sigma_j) \geq \nu_2$. Informally, the m-OGP means that one cannot find m near-optimal solutions for m instances such that all $m(m-1)/2$ pairs of distances lie in the interval (ν_1, ν_2) . Clearly, the case $m=2$ boils down to the e-OGP discussed earlier. In many applications, the difference between ν_1 and ν_2 is often significantly smaller than the value of ν_1 itself; that is, $\nu_1 \approx \nu_2$. Thus, roughly speaking, the m-OGP means that one cannot find m near-optimal solutions so that all pairwise distances are $\approx \nu_1$. The first usage of OGP as an obstruction to algorithms was adopted by Gamarnik and Sudan in ref. 46. The first application of m-OGP as an obstruction to algorithms was adopted by Rahman and Virag in ref. 47. While their variant of m-OGP proved useful for some applications (29), since then, the definition of m-OGP has been extended to "less symmetric" variants, as one by Wein in ref. 48 and by Bresler and Huang in ref. 30. We will not discuss the nature of these extensions and, instead, refer to the aforementioned references for further details. e-OGP was introduced by Chen et al. (49).

OGP Is an Obstruction to Stability. Next, we discuss how the presence of variants of the OGP constitutes an obstruction to algorithms. The broadest class of algorithms for which such an implication can be established is algorithms that are stable with respect to a small perturbation of instances ξ . An algorithm \mathcal{A} is viewed here simply as a mapping from instances ξ to solutions σ , which we abbreviate as $\mathcal{A}(\xi) = \sigma$. Suppose we have a parametrized

collection of instances ξ_t with discrete parameter t taking values in some interval $[0, T]$. We say that the algorithm \mathcal{A} is stable, or specifically κ -stable, with respect to the family (ξ_t) if for every t , $\rho_N(\mathcal{A}(\xi_{t+1}), \mathcal{A}(\xi_t)) \leq \kappa$. Informally, if we think of ξ_{t+1} as an instance obtained from ξ_t by a small perturbation, the output of the algorithm does not change much: The algorithm is not very sensitive to the input. Continuous versions of such stability have been considered as well, specifically in the context of models with Gaussian distribution. Here, t is a continuous parameter, and the algorithm is stable with sensitivity value δ if for every t $\rho_N(\mathcal{A}(\xi_{t+\delta}), \mathcal{A}(\xi_t)) \leq \kappa$. Often, these bounds are established only with probabilistic guarantees, both in the case of discrete and continuously valued t .

Now, assume that the e-OGP holds for a family of instances $\xi_t, t \in [0, T]$. Assume two additional conditions: 1) The regime $\rho_N(\sigma, \tau) \leq \nu_1$ for μ -optimal solutions σ of ξ_0 , and μ -optimal solutions τ of ξ_T is nonexistent (namely, every μ -optimal solution of ξ_0 is far from every μ -optimal solution of ξ_T); and 2) $\kappa < \nu_2 - \nu_1$. The property (1) is typically verified since the endpoints ξ_0 and ξ_T of the interpolated sequence are often independent, and (1) can be checked by straightforward moment arguments. The verification of condition (2) typically involves technical and problem-dependent arguments. The conditions (1) and (2) above plus the OGP allow us to conclude that any κ -stable algorithm fails to produce a μ -optimal solution. This is seen by a simple contradiction argument based on continuity: Consider the sequence of solutions $\sigma_t = \mathcal{A}(\xi_t), t = 0, \dots, T$, produced by the algorithm \mathcal{A} . We have $\rho_N(\sigma_t, \sigma_{t+1}) \leq \kappa$ for every t and $\rho_N(\sigma_0, \sigma_T) \geq \nu_2$ by the OGP and assumption (1) above. Suppose, for the purposes of contradiction, that every solution σ_t produced by the algorithm is μ -optimal for every instance ξ_t . Then, by the OGP, it is the case that, in particular, $\rho_N(\sigma_0, \sigma_t)$ is either at most ν_1 or at least ν_2 for every t . Since $\rho_N(\sigma_0, \sigma_T) \geq \nu_2$, there exists t , possibly $t=0$, such that $\rho_N(\sigma_0, \sigma_t) \leq \nu_1$ and $\rho_N(\sigma_0, \sigma_{t+1}) \geq \nu_2$. Then, $\kappa \geq \rho_N(\sigma_t, \sigma_{t+1}) \geq |\rho_N(\sigma_0, \sigma_t) - \rho_N(\sigma_0, \sigma_{t+1})| \geq \nu_2 - \nu_1$, which is a contradiction with the assumption (2).

Simply put, the crux of the argument is that the algorithm cannot "jump" over the gap $\nu_2 - \nu_1$, since the incremental distances, bounded by κ , are too small to allow for it. This is the main method by which stable algorithms are ruled out in the presence of the OGP and is illustrated in *SI Appendix, Fig. S4*. In this figure, the smaller circle represents all of the μ -optimal solutions that are at most distance ν_1 from σ_0 , across all instances $t \in [0, T]$. The complement to the larger circle represents all of the μ -optimal solutions that are at least distance ν_2 from σ_0 , again across all instances $t \in [0, T]$. All μ -optimal solutions across all instances should fall into one of these two regions, according to the e-OGP. As the sequence of solutions $\mathcal{A}(\xi_t), t \in [0, T]$ has to cross between these two regions, at some point t , the distance between $\mathcal{A}(\xi_t)$ and $\mathcal{A}(\xi_{t+1})$ will have to exceed κ , contradicting the stability property.

Suppose the model does not exhibit the OGP, but does exhibit the m-OGP (ensemble version). We now describe how this can also be used as an obstruction to stable algorithms. As above, the argument is based on first verifying that for two independent instances ξ and $\tilde{\xi}$, the regime $\rho_N(\sigma, \tau) \leq \nu_1$ is not possible with overwhelming probability. Consider $m + 1$ independent instances $\xi_0, \xi_1, \dots, \xi_m$ and consider an interpolated sequence $\xi_{1,t}, \dots, \xi_{m,t}, t \in [0, T]$ with the property that $\xi_{i,T} = \xi_i$ and $\xi_{i,0} = \xi_0$ for all $i = 1, 2, \dots, m$. In other words, the sequence starts with m identical copies of instance ξ_0 and slowly interpolates toward m instance ξ_1, \dots, ξ_m . The details of such a construction are usually guided by concrete problems. Typically, such constructions are also symmetric in distribution, so that, in particular, all pairwise expected distances between the

algorithmic outputs $\mathbb{E}[\rho_N(\mathcal{A}(\xi_{i,t}), \mathcal{A}(\xi_{j,t}))]$ are the same, say, denoted by $\Delta_t, t \in [0, T]$. Furthermore, in some cases, a concentration around the expectation can be also established. As an implication, this set of identical pairwise distances Δ_t spans values from zero to value larger than ν_1 by the assumption (1). But the stability of the algorithm \mathcal{A} also implies that at some point t , it will be the case that $\Delta_t \in (\nu_1, \nu_2)$, contradicting the m-OGP. This is illustrated in *SI Appendix, Fig. S5*. The high-concentration property discussed above was established by using standard methods in refs. 29 and 48, but the approach in ref. 50 was based on different arguments employing methods from the Ramsey extremal combinatorics field. Here, the idea is to generate many more $M \gg m$ independent instances than the value of m arising in m-OGP and showing using Ramsey theory results that there should exist a subset of m -instances such that all pairwise distances fall into interval (ν_1, ν_2) .

OGP for Concrete Models. We now illustrate the OGP and its variants for the Number Partitioning problem and the maximum clique problem as examples. The discussion will be informal, and references containing detailed derivations will be provided. The instance ξ of the Number Partitioning problem is a sequence X_1, \dots, X_N of independent standard normal random variables. For every binary vector $\sigma \in \{-1, 1\}^N$, the inner product $Z(\sigma) = \langle \xi, \sigma \rangle$ is a mean zero normal random variable with variance N . The likelihood that $Z(\sigma)$ takes value in an interval around zero with length $\sqrt{N}2^{-N}$ is roughly $(2\pi)^{-\frac{1}{2}}$ times the length of the interval, namely, times 2^{-N} . Thus, the expected number of such vectors σ is of constant order $(2\pi)^{-\frac{1}{2}}$ since there are 2^N choices for σ . This gives an intuition for why there do exist vectors achieving value $\sqrt{N}2^{-N}$. Details can be found in ref. 11. Any choice of order magnitude smaller length interval leads to expectation converging to zero and, thus, is nonachievable with overwhelming probability. Now, to discuss the OGP, fix constants $\alpha, \rho \in (0, 1)$ and consider pairs of vectors σ, τ , which achieve objective value $\sqrt{N}2^{-\alpha N}$ and have the scaled inner product $N^{-1} \langle \sigma, \tau \rangle = \rho$. The expected number of such pairs is roughly $\binom{N}{\frac{1-\rho}{2}N}$ times the area of the square with side length approximately $2^{-\alpha N}$, namely, $2^{-2\alpha N}$. This is because the likelihood that a pair of correlated Gaussians $Z(\sigma), Z(\tau)$ falls into a very small square around zero is dominated by the area of this square up to a constant. Approximating $\binom{N}{\frac{1-\rho}{2}N}$ as $\exp(NH(\frac{1-\rho}{2}))$, where $H(x) = -x \log x - (1-x) \log(1-x)$ is the standard binary entropy, the expectation of the number of such pairs evaluates to roughly $\exp(N(H(\frac{1-\rho}{2}) - 2\alpha \log 2))$. As the largest value of $H(\cdot)$ is $\log 2$, we obtain an important discovery: For every $\alpha \in (1/2, 1)$, there is a region of ρ of the form $(\rho_0, 1)$ for which the expectation is exponentially small in N , and, thus, the value $\sqrt{N}2^{-\alpha N}$ is not achievable by pairs of partitions σ and τ with scaled inner products in $(\rho_0, 1)$, with overwhelming probability. Specifically, for every pair of μ -optimal solutions σ, τ with $\mu = (2^{-\alpha N} - 2^{-N})$, it is the case that $N^{-1} \langle \sigma, \tau \rangle$ is either one or at most ρ_0 . With this value of μ and $\nu_1 = 0, \nu_2 = N(1 - \rho_0)/2$, we conclude that the model exhibits the OGP.

Showing the ensemble version of OGP (namely, e-OGP) can be done using very similar calculations, but what about values of α smaller than $1/2$? After all, the state-of-the-art algorithms achieve only order $2^{-O(\log^2 N)}$, which corresponds to $\alpha = 0$ at scale. It turns out (50) that, in fact, the m-OGP holds for this model for all strictly positive α with a judicious choice of constant value m . Furthermore, the m-OGP extends all the way to $2^{-O(\sqrt{N \log N})}$ by taking m

growing with N , but not beyond that value, at least using the available proof techniques. Thus, at this stage, we may conjecture that the problem is algorithmically hard for objective values smaller than order $2^{-O(\sqrt{N \log N})}$, but beyond this value, we don't have plausible predictions either for hardness or tractability.

The largest clique problem is another example where the OGP-based predictions are consistent with the state-of-the-art algorithms. Recall from the earlier section that while the largest clique in $\mathbb{G}(N, 1/2)$ is roughly $2 \log_2 N$, the largest size clique achievable by known algorithm is roughly half-optimal with size approximately $\log_2 N$. Fixing $\alpha \in (1, 2)$, consider pairs of cliques σ, τ with sizes roughly $\alpha \log_2 N$. This range of α corresponds to clique sizes known to be present in the graph existentially with overwhelming probability, but not achievable algorithmically. As it turns out (46, 51), the model exhibits the OGP for $\alpha \in (1 + 1/\sqrt{2}, 2)$ and overlap parameters $\nu_1(\alpha) < \nu_2(\alpha)$, both of which are at the scale $\log_2 N$. Specifically, every two cliques of size $\alpha \log_2 N$ have intersection size either at least $\tilde{\nu}_1(\alpha)$ or at most $\tilde{\nu}_2(\alpha)$, for some values $\tilde{\nu}_1(\alpha) > \tilde{\nu}_2(\alpha)$, both at scale $\log_2 n$. This easily implies OGP with $\nu_j(\alpha) = 2\alpha - \tilde{\nu}_j(\alpha), j = 1, 2$. The calculations in refs. 46 and 51 were carried out for a similar problem of finding a largest independent set (a set with no internal edges) in sparse random graphs $G(N, p_N)$ with p_N of the order $O(1/N)$, but the calculation details for the dense graph $\mathbb{G}(N, 1/2)$ are almost identical. The end result is that $\tilde{\nu}_j(\alpha) = x_j(\alpha) \log_2 N$, where $x_1(\alpha)$ and $x_2(\alpha)$ are the two roots of the quadratic equation $x^2/2 - x + 2\alpha - \alpha^2 = 0$, namely, $1 \pm \sqrt{1 - 2(2\alpha - \alpha^2)}$, and roots exist if and only if $\alpha > 1 + 1/\sqrt{2}$. For recent algorithmic results for the related problem on random regular graphs with small degree, see ref. 52.

To illustrate the ensemble version of the OGP, consider an ensemble $G_1, \dots, G_{\binom{N}{2}}$ of random $\mathbb{G}(N, 1/2)$ graphs, constructed as follows. Generate G_0 according to the probability law of $\mathbb{G}(N, 1/2)$. Introduce a random uniform order π of $\binom{N}{2}$ pairs of nodes $1, \dots, N$, and generate graph G_t from G_{t-1} by resampling the pair (i_t, j_t) —the t -th pair in the order π —and leaving other edges of G_{t-1} intact. This way, each graph G_t individually has the law of $\mathbb{G}(N, 1/2)$, but neighboring graphs G_{t-1} and G_t differ by at most one edge. The beginning and the end graphs G_0 and $G_{\binom{N}{2}}$ are clearly independent. Now, fix $\alpha \in (1 + 1/\sqrt{2}, 2)$ and $x \in (0, \alpha)$. A simple calculation shows that the expected number of $\alpha \log_2 N$ size cliques σ in G_{t_1} and same-size cliques τ in G_{t_2} , such that the size of the intersection of σ and τ is $x \log_2 N$, is approximately given by $\exp((1 + o(1)) \log_2^2 N ((1 - \rho)x^2/2 - x + 2\alpha - \alpha^2))$, where $\rho = (t_2 - t_1)/\binom{N}{2}$ is the (rescaled) number of edges of G_{t_2} , which were resampled from G_{t_1} . Easy calculation shows that the function $(1 - \rho)x^2/2 - x + 2\alpha - \alpha^2$ has two roots $x_1 > x_2$ in the interval $(0, \alpha)$ given by $(1 - \rho)^{-1} \left(1 \pm \sqrt{1 - 2(1 - \rho)(2\alpha - \alpha^2)} \right)$, when the larger root is smaller than α , namely, when $\rho < \frac{2}{\alpha} - 1$. On the other hand, when $\rho > \frac{2}{\alpha} - 1$, including the extreme case $\rho = 1$ corresponding to $t_1 = 0, t_2 = \binom{N}{2}$, there is only one root in $(0, \alpha)$. The model thus not only does exhibit the e-OGP for the ensemble $G_t, 0 \leq t \leq \binom{N}{2}$. Additionally, the value $\rho^* = \frac{2}{\alpha} - 1$ represents a new type of phase transition. When $\rho < \rho^*$, every two cliques with size $\alpha \log_2 N$ have intersection size either at least $x_1 \log_2 N$ or at most $x_2 \log_2 N$. On the other hand, when $\rho > \rho^*$, the second regime disappears, and any two such cliques have intersection size only at most $x_2 \log_2 N$. The situation is illustrated in *SI Appendix, Fig.*

S6 for the special case $\alpha = 1.72$ and $\rho^* = 2/\alpha - 1 = 0.163$.

The value $\alpha = 1 + 1/\sqrt{2}$ is not tight with respect to the apparent algorithmic hardness, which occurs at $\alpha = 1$, as discussed earlier. This is where the multi-OGP comes to rescue. A variant of an (asymmetric) m-OGP has been established recently in ref. 48, building on an earlier symmetric version of m-OGP, discovered in ref. 47, ruling out the class of stable algorithms as potential contenders for solving the Independent Set problem above $\alpha = 1$. More specifically, in both works, it is shown that for every fixed $\epsilon > 0$, there exists a constant $m = m(\epsilon)$ such that the model exhibits m-OGP with this value of m for independent sets corresponding to $\alpha = 1 + \epsilon$.

Stability of Low-Degree Polynomials. How does one establish the link of the form OGP implies Class \mathcal{C} of algorithms fails, and what classes \mathcal{C} of algorithms can be proven to fail in presence of the OGP? As discussed earlier, the failure is established by showing stability (insensitivity) of the algorithms in class \mathcal{C} to the changes of the input. The largest class of algorithms to the day that is shown to be stable is the class of algorithms informally dubbed as "low-degree polynomials." This class of algorithms has deep connections with the so-called Sum-of-Squares method (42–44). Roughly speaking, this is a class of algorithms where the solution $\sigma \in \Sigma_N$ is obtained by constructing an N -sequence of multivariate polynomials $p_j(\xi), j = 1, \dots, N$ with variables evaluated at entries of the instance $\xi \in \Xi_N$. The largest degree d of the polynomial is supposed to be low, though depending on the problem, it can take even significantly high value. What can be said about this method in the context of the problems exhibiting the OGP? It was shown in refs. 30, 48, and 51 that degree- d polynomial-based algorithms are stable. This is true even when d is as large as $O(N/\log N)$. Thus, algorithms based on polynomials even with degree $O(N/\log N)$ fail to find near-optimum solutions for these problems due to the presence of the OGP. Many algorithms can be modeled as special cases of low-degree polynomials, and therefore ruled out by OGP, including the so-called Approximate Message Passing algorithms (51, 53), local algorithms considered in refs. 46 and 47, and quantum versions of local algorithms known as QAOA (4).

OGP and the Problem of Finding Ground States of p -Spin Models. There is another large class of optimization problems for which the OGP approach provides a tight classification of solvable vs. not yet solvable in polynomial time problems. It is the problem of finding ground states of p -spin models—the very class of optimization problems that led to studying the overlap distribution and overlap gap-type properties to begin with. The problem is described as follows: Suppose one is given a p -order tensor J of side length N . Namely, J is described as a collection of real values J_{i_1, \dots, i_p} , where i_1, \dots, i_p range between one and N . Given the solution space Σ_N , the goal is finding a solution $\sigma \in \Sigma_N$ (called ground state), which minimizes the inner product $\langle J, \sigma \rangle = \sum_{i_1, \dots, i_p} J_{i_1, \dots, i_p} \sigma_{i_1} \dots \sigma_{i_p}$. A common assumption, capturing most of the structural and algorithmic complexity of the model, is that the entries J_{i_1, \dots, i_p} are independent standard normal entries. Two important special cases of the solution space Σ_N include the case when Σ_N is a binary cube $\{-1, 1\}^N$ and when it is a sphere of a fixed radius (say, one) in the N -dimensional real space, the latter case being referred to as the *spherical* spin-glass model. The special case $p = 2$ and $\Sigma_N = \{-1, 1\}^N$ is the celebrated

Sherrington–Kirkpatrick model (54)—the centerpiece of the spin-glass theory. The problem of computing the value of this optimization problem above was at the core of the spin-glass theory in the past four decades, which led to the replica symmetry-breaking technique first introduced in the physics literature by Parisi (25) and then followed by a series of mathematically rigorous developments, including those by Talagrand (55) and Panchenko (56).

The algorithmic progress on this problem, however, is quite recent. In the important development by Subag (57) and Montanari (58), polynomial time algorithms were constructed for finding near ground states in these models. Remarkably, this was done in the regimes where the model does not exhibit the OGP. A follow-up development in ref. 59 extended this to models that do exhibit OGP, but in this case, near ground states are not reached. Instead, one constructs the best solution within the part of the solution space not exhibiting the OGP. To put it differently, the algorithms produce a μ -optimal solution for the smallest value of μ for which the OGP does not hold. Conversely (51, 53, 60), the optimization problem above cannot be solved within the class of stable algorithms.

OGP, the Clustering Property, and the Curious Case of the Perceptron Model.

We now return to the subject of weak and strong clustering property and discuss it from the perspective of the OGP and algorithmic hardness. The relationship is nontrivial and perhaps best exemplified by yet another model, which exhibits statistical to computation gap: the binary perceptron model. The input to the model is $M \times N$ matrix of i.i.d. standard normal entries X_{ij} , $1 \leq i \leq M$, $1 \leq j \leq N$. A fixed parameter $\kappa > 0$ is fixed. The goal is to find a binary vector $\sigma \in \{\pm 1\}^N$ such that for each i , $\sum_{1 \leq j \leq N} X_{ij} \sigma_j \in [-\kappa, \kappa]$. This is also known as the binary symmetric perceptron model, to contrast it with the asymmetric model, where the requirement is that $X\sigma \geq \kappa$, for some fixed parameter κ (positive or negative), with inequalities interpreted coordinate-wise. In the asymmetric case, a typical choice is $\kappa = 0$. Replica symmetry-breaking-based heuristic methods developed in this context by Krauth and Mézard (61) predict a sharp threshold at $\alpha_{\text{SAT}} = 0.83\dots$ for the case $\kappa = 0$, so that when $\alpha < \alpha_{\text{SAT}}$ and $M \leq \alpha N$ such a vector σ exists, and when $\alpha > \alpha_{\text{SAT}}$ such a vector σ does not exist, both with high probability as $N \rightarrow \infty$. This was confirmed rigorously as a lower bound in ref. 62. For the symmetric case, a similar sharp threshold is known at a κ -dependent constant $\alpha_{\text{SAT}}(\kappa)$, as established in refs. 63 and 64.

The known algorithmic results, however, are significantly weaker. The algorithm by Kim and Rouche (65) finds σ for the asymmetric $\kappa = 0$ case when $\alpha < 0.005$, notably much smaller than α_{SAT} . It is quite likely (though not known) that a version of this algorithm should work in the symmetric case as well for some positive κ -dependent constant α . Heuristically, the message-passing algorithm was found to be effective at small positive densities, as reported in ref. 66. Curiously, though, it is known that the symmetric model exhibits the weak clustering property at all positive densities $\alpha > 0$ (63, 64)! Furthermore, quite remarkably, each cluster consists of singletons! Namely, the cardinality of each cluster is one, akin to the “needles in the haystack” metaphor. A similar picture was established in the context of the Number Partitioning problem (50). This interesting entropic (due to the subextensive cardinality of clusters) phenomenon is fairly novel, and its algorithmic implications have been also discussed in ref. 67. Thus, assuming the extendibility of the Kim–Roche algorithm and/or validity of the message-passing algorithm, we have an example of a

model exhibiting weak clustering property, but amenable to polynomial time algorithms. As an explanation of this phenomena, the aforementioned references point out the “nonuniformity” in algorithmic choices of the solutions, allowing it somehow to bypass the overwhelming clustering picture exhibited by the weak clustering property.

What about the strong clustering property and the OGP? For any model, the presence of the pairwise ($m = 2$) OGP for a single instance implies the strong clustering property: If every pair of μ -optimal solutions is at distance at most ν_1 or at least $\nu_2 > \nu_1$, and the second case is nonvacuous, clearly, the set of solutions is strongly clustered. It, furthermore, indicates that the diameter of each cluster is strictly smaller (at scale) than the distances between the clusters (SI Appendix, Figs. S2 and S3). In fact, this is precisely how the strong clustering property was discovered to begin with in refs. 27 and 28. However, the converse does not need to be the case, as one could not rule out an example of a strong clustering property where the cluster diameters are larger than the distances between them, and, as a result, the overlaps of pairs of solutions span continuous intervals. At this stage, we are not aware of any such examples.

Back to the symmetric perceptron model, it is known that the model does exhibit the OGP, at a value α strictly smaller than the critical threshold $\alpha_{\text{SAT}}(\kappa)$ (68), but strictly larger than the algorithmic value 0.005 achieved algorithmically in ref. 65. As a result, the strong clustering property holds as well above this α . Naturally, we conjecture that the problem of finding a solution σ is hard in this regime. In conclusion, the binary perceptron model demonstrates that the weak clustering property is provably not an obstruction to polynomial time algorithms, but OGP and its implication, the strong clustering property, likely is, at least for the case of stable algorithms.

Finally, as the strong clustering property appears to be the most closely related to the OGP, it raises a question as to whether it can be used as a “signature” of hardness in lieu of OGP and/or whether it can be used as a method to rule out classes of algorithm, similarly to the OGP-based method. There are two challenges associated with this proposal. First, we are not aware of any effective method of establishing the strong clustering property other than the one based on the OGP. Second, and more importantly, it is not entirely clear how to link meaningfully the strong clustering property with m-OGP, which appears necessary for many problems in order to bridge the computational gaps to the known algorithmically achievable values. Likewise, it is not entirely clear what is the appropriate definition of either weak or strong clustering property in the ensemble version, when one considers a sequence of correlated instances, which is again another important ingredient in the implementation of the refutation-type arguments. Perhaps some form of dynamically evolving clustering picture is of relevance here. Some recent progress on the former question regarding the geometry of the m-OGP was obtained recently in ref. 69 in the context of the spherical spin-glass model. We leave both questions as interesting open venues for future investigation.

Discussion

In this article, we have discussed a new approach to computational complexity arising in optimality studies of random structures. The approach, which is based on topological disconnectedness of the overlaps of near-optimal solutions, called the OGP, is both a theory and a method. As a theory, it predicts the onset of algorithmic hardness in random structures at the onset of the OGP. The approach has been verified at various levels of precision in many

classes of problems, and the summary of the state of the art is presented in *SI Appendix, Table S1*. In this table, the current list of models known to exhibit an apparent algorithmic hardness is provided, along with references and notes indicating whether the OGP-based method matches the state-of-the-art algorithmic knowledge. The “Yes” indicates this to be the case, and “Not known” indicates that the formal analysis has not been completed yet. Remarkably, to the day, we are not aware of a model that does not exhibit the OGP, but that is known to exhibit some form of an apparent algorithmic hardness.

The OGP-based approach also provides a concrete method for rigorously establishing barriers for classes of algorithms. Typically, such barriers are established by verifying certain stability (input insensitivity) of algorithms, making them inherently incapable of

overcoming the gaps appearing in the overlap structures. While the general approach for ruling out such classes of algorithm is more or less problem-independent, the exact nature of such stability, as well as the OGP itself, is very much problem-dependent, and the required mathematical analysis varies from borderline trivial to extremely technical, often relying on the state-of-the-art development in the field of mathematical theory of spin glasses.

Data Availability. There are no data underlying this work.

Acknowledgments

This work was supported by NSF Grant DMS-2015 517. I thank the anonymous reviewers for many suggestions, which led to significant improvement of the presentations of the manuscript.

- 1 R. Vershynin, *High-Dimensional Probability: An Introduction with Applications in Data Science* (Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, UK, 2018), vol. 47.
- 2 P. Bühlmann, S. Van De Geer, *Statistics for High-Dimensional Data: Methods, Theory and Applications* (Springer Series in Statistics, Springer Science & Business Media, Berlin, Germany, 2011).
- 3 S. Foucart, H. Rauhut, *A Mathematical Introduction to Compressive Sensing* (Applied and Numerical Harmonic Analysis, Birkhäuser, New York, 2013).
- 4 E. Farhi, D. Gamarnik, S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: A typical case. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2004.09002> (Accessed 15 September 2021).
- 5 E. Farhi, D. Gamarnik, S. Gutmann, The quantum approximate optimization algorithm needs to see the whole graph: Worst case examples. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2005.08747> (Accessed 15 September 2021).
- 6 E. Farhi, J. Goldstone, S. Gutmann, A quantum approximate optimization algorithm. *arXiv [Preprint]* (2014). <https://arxiv.org/abs/1411.4028> (Accessed 15 September 2021).
- 7 N. Alon, J. H. Spencer, *The Probabilistic Method* (John Wiley & Sons, New York, 2004).
- 8 R. M. Karp, “The probabilistic analysis of some combinatorial search algorithms” in *Algorithms and Complexity: New Directions and Recent Results*, J. F. Traub, Ed. (Academic, New York, 1976), vol. 1, pp. 1–19.
- 9 M. Jerrum, Large cliques elude the metropolis process. *Random Structures Algorithms* **3**, 347–359 (1992).
- 10 M. R. Garey, D.S. Johnson, *Computers and Intractability* (Series of Books in the Mathematical Sciences, Freeman, San Francisco, CA, 1979), vol. 174.
- 11 N. Karmarkar, R. M. Karp, *The Differencing Method of Set Partitioning* (Computer Science Division, University of California Berkeley, Berkeley, CA, 1982).
- 12 S. Boettcher, S. Mertens, Analysis of the Karmarkar-Karp differencing algorithm. *Eur. Phys. J. B* **65**, 131–140 (2008).
- 13 J. Hastad, Clique is hard to approximate within n^{1-c} . *Acta Math.* **182**, 105–142 (1999).
- 14 R. Lipton, New directions in testing. *Distributed Comput. Cryptogr.* **2**, 191–202 (1991).
- 15 D. Gamarnik, E. C. Kızıldağ, Computing the partition function of the Sherrington–Kirkpatrick model is hard on average. *Ann. Appl. Probab.* **31**, 1474–1504 (2021).
- 16 M. Ajtai, “Generating hard instances of lattice problems” in *Proceedings of the Twenty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 1996), pp. 99–108.
- 17 A. Choromanska, M. Henaff, M. Mathieu, G. Ben Arous, Y. LeCun, “The loss surfaces of multilayer networks” in *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics (AISTATS)* (JMLR, 2015), pp. 192–204.
- 18 Y. Fu, P. W. Anderson, Application of statistical mechanics to NP-complete problems in combinatorial optimisation. *J. Phys. Math. Gen.* **19**, 1605 (1986).
- 19 S. Kirkpatrick, B. Selman, Critical behavior in the satisfiability of random Boolean expressions. *Science* **264**, 1297–1301 (1994).
- 20 R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, L. Troyansky, Determining computational complexity from characteristic ‘phase transitions’. *Nature* **400**, 133–137 (1999).
- 21 M. Mezard, G. Parisi, M. A. Virasoro, *Spin-Glass Theory and Beyond* (Lecture Notes in Physics, World Scientific, Singapore, 1987), vol. 9.
- 22 M. Mezard, A. Montanari, *Information, Physics and Computation* (Oxford Graduate Texts, Oxford University Press, Oxford, UK, 2009).
- 23 J. Ding, A. Sly, N. Sun, “Proof of the satisfiability conjecture for large k ” in *STOC’15: Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2015), pp. 59–68.
- 24 E. Friedgut, Sharp thresholds of graph properties, and the k -SAT problem. *J. Am. Math. Soc.* **4**, 1017–1054 (1999).
- 25 G. Parisi, A sequence of approximated solutions to the SK model for spin glasses. *J. Phys. Math. Gen.* **13**, L115 (1980).
- 26 A. Coja-Oghlan, A better algorithm for random k -SAT. *SIAM J. Comput.* **39**, 2823–2864 (2010).
- 27 D. Achlioptas, F. Ricci-Tersenghi, “On the solution-space geometry of random constraint satisfaction problems” in *STOC’06: Proceedings of the Thirty-Eighth Annual ACM Symposium on Theory of Computing* (Association for Computing Machinery, New York, NY, 2006), pp. 130–139.
- 28 M. Mézard, T. Mora, R. Zecchina, Clustering of solutions in the random satisfiability problem. *Phys. Rev. Lett.* **94**, 197205 (2005).
- 29 D. Gamarnik, M. Sudan, Performance of sequential local algorithms for the random NAE- k -SAT problem. *SIAM J. Comput.* **46**, 590–619 (2017).
- 30 G. Bresler, B. Huang, The algorithmic phase transition of random k -sat for low degree polynomials. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2106.02129> (Accessed 15 September 2021).
- 31 S. Hetterich, “Analysing survey propagation guided decimation random formulas” in *43rd International Colloquium on Automata, Languages, and Programming (ICALP 2016)*, I. Chatzigiannakis, M. Mitzenmacher, Y. Rabani, D. Sangiorgi, Eds. (Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, Dagstuhl, Germany, 2016), pp. 65:1–65:12.
- 32 A. Coja-Oghlan, A. Haqshenas, S. Hetterich, Walksat stalls well below satisfiability. *SIAM J. Discrete Math.* **31**, 1160–1173 (2017).
- 33 M. Mézard, G. Parisi, R. Zecchina, Analytic and algorithmic solution of random satisfiability problems. *Science* **297**, 812–815 (2002).
- 34 C. P. Gomes, B. Selman, Computer science. Satisfied with physics. *Science* **297**, 784–785 (2002).
- 35 F. Ricci-Tersenghi, G. Semerjian, On the cavity method for decimated random constraint satisfaction problems and the analysis of belief propagation guided decimation algorithms. *J. Stat. Mech.* **2009**, P09001 (2009).
- 36 A. Coja-Oghlan, Belief propagation guided decimation fails on random formulas. *J. ACM*, **63**, 1–55 (2017).
- 37 F. Ricci-Tersenghi, Mathematics. Being glassy without being hard to solve. *Science* **330**, 1639–1640 (2010).
- 38 L. Budzynski, F. Ricci-Tersenghi, G. Semerjian, Biased landscapes for random constraint satisfaction problems. *J. Stat. Mech.* **2019**, 023302 (2019).
- 39 L. Budzynski, G. Semerjian, Biased measures for random constraint satisfaction problems: larger interaction range and asymptotic expansion. *J. Stat. Mech.* **2020**, 103406 (2020).

- 40 F. Krzakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, L. Zdeborová, Gibbs states and the set of solutions of random constraint satisfaction problems. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10318–10323 (2007).
- 41 R. Marino, G. Parisi, F. Ricci-Tersenghi, The backtracking survey propagation algorithm for solving random K-SAT problems. *Nat. Commun.* **7**, 12996 (2016).
- 42 S. B. Hopkins, D. Steurer, “Efficient Bayesian estimation from few samples: community detection and related problems” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2017), pp. 379–390.
- 43 S. B. Hopkins et al., “The power of sum-of-squares for detecting hidden structures” in *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, New York, 2017), pp. 720–731.
- 44 S. Hopkins, “Statistical inference and the sum of squares method,” PhD thesis, Cornell University, Ithaca, NY (2018).
- 45 D. Gamarnik, Q. Li, Finding a large submatrix of a Gaussian random matrix. *Ann. Stat.* **46** (6A), 2511–2561 (2018).
- 46 D. Gamarnik, M. Sudan, Limits of local algorithms over sparse random graphs. *Ann. Probab.* **45**, 2353–2376 (2017).
- 47 M. Rahman, B. Virag, Local algorithms for independent sets are half-optimal. *Ann. Probab.* **45**, 1543–1577 (2017).
- 48 A. S. Wein, Optimal low-degree hardness of maximum independent set. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2010.06563> (Accessed 15 September 2021).
- 49 W.-K. Chen, D. Gamarnik, D. Panchenko, M. Rahman, Suboptimality of local algorithms for a class of max-cut problems. *Ann. Probab.* **47**, 1587–1618 (2019).
- 50 D. Gamarnik, E. C. Kızıldağ, Algorithmic obstructions in the random number partitioning problem. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2103.01369> (Accessed 15 September 2021).
- 51 D. Gamarnik, A. Jagannath, A. S. Wein, “Low-degree hardness of random optimization problems” in *61st Annual Symposium on Foundations of Computer Science* (2020).
- 52 R. Marino, S. Kirkpatrick, Large independent sets on random d -regular graphs with d small. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2003.12293> (Accessed 15 September 2021).
- 53 D. Gamarnik, A. Jagannath, The overlap gap property and approximate message passing algorithms for p -spin models. *Ann. Probab.* **49**, 180–205 (2021).
- 54 D. Sherrington, S. Kirkpatrick, Solvable model of a spin-glass. *Phys. Rev. Lett.* **35**, 1792 (1975).
- 55 M. Talagrand, *Basic Examples* (Mean Field Models for Spin Glasses, Springer, Berlin, Germany, 2010), vol. 1.
- 56 D. Panchenko, *The Sherrington-Kirkpatrick Model* (Springer Monographs in Mathematics, Springer Science & Business Media, New York, 2013).
- 57 E. Subag, Following the ground states of full-RSB spherical spin glasses. *Commun. Pure Appl. Math.* **74.5**, 1021–1044 (2021).
- 58 A. Montanari, Optimization of the Sherrington–Kirkpatrick Hamiltonian. *SIAM J. Sci. Comput.* **2021**, FOCS19-1 (2021).
- 59 A. El Alaoui, A. Montanari, M. Sellke, Optimization of mean-field spin glasses. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2001.00904> (Accessed 15 September 2021).
- 60 M. Sellke, Approximate ground states of hypercube spin glasses are near corners. *arXiv [Preprint]* (2020). <https://arxiv.org/abs/2009.09316> (Accessed 15 September 2021).
- 61 W. Krauth, M. Mézard, Storage capacity of memory networks with binary couplings. *J. Phys. (Paris)* **50**, 3057–3066 (1989).
- 62 J. Ding, N. Sun, “Capacity lower bound for the Ising perceptron” in *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2019), pp. 816–827.
- 63 W. Perkins, C. Xu, “Frozen 1-RSB structure of the symmetric Ising perceptron” in *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (Association for Computing Machinery, New York, 2021), pp. 1579–1588.
- 64 E. Abbe, S. Li, A. Sly, Proof of the contiguity conjecture and lognormal limit for the symmetric perceptron. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2102.13069> (Accessed 15 September 2021).
- 65 J. H. Kim, J.R. Roche, Covering cubes by random half cubes, with applications to binary neural networks. *J. Comput. Syst. Sci.* **56**, 223–252 (1998).
- 66 A. Braunstein, R. Zecchina, Learning by message passing in networks of discrete synapses. *Phys. Rev. Lett.* **96**, 030201 (2006).
- 67 M. Bellitti, F. Ricci-Tersenghi, A. Scardicchio, Entropic barriers as a reason for hardness in both classical and quantum algorithms. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2102.00182> (Accessed 15 September 2021).
- 68 C. Baldassi, R. Della Vecchia, C. Lucibello, R. Zecchina, Clustering of solutions in the symmetric binary perceptron. *J. Stat. Mech.* **2020**, 073303 (2020).
- 69 G. Ben Arous, A. Jagannath, Shattering versus metastability in spin glasses. *arXiv [Preprint]* (2021). <https://arxiv.org/abs/2104.08299> (Accessed 15 September 2021).